DOCUMENT RESUME

ED 147 352 TM 006 749

AUTHOR	Myers, Charles T.
TITLE	Bias and Interpretation: Cases for Ordinal Measurement.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RM-73-18
PUB DATE	Sep 73
NOTE	10p.; Paper presented at the Annual Meeting of the American Psychological Association (81st, Montreal, Quebec, Canada, August 27-31, 1973)
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	*Analysis of Covariance; College Students; Equated Scores; *Grade Point Average; Grades (Scholastic); Higher Education; High School Students; *Mathematical Models; Pass Fail Grading; *Scores; Senior High Schools; Standardized Tests; Statistical Analysis; Student Distribution; *Test Bias; *Test Interpretation; Test Reliability
IDENTIFIERS	Equipercentile Equating Method; *Ordinal Scales

ABSTRACT

High school and college grades are suggested to have the nature of ordinal scales rather than interval scales; hence the median is the appropriate and preferred average, rather than the mean. The use of a median grade point average has some of the major advantages of pass-fail grading. Assuming grades to be ordinal rather than interval data suggests that estimates of test bias should not be subjected only to analysis of covariance. As an alternative, equipercentile equating procedures are suggested as a method of studying bias that would tend to be more favorable for low-scoring groups. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED147352

RESEARCH MEMORANDUM

RM-73-18

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BIAS AND INTERPRETATION: CASES FOR ORDINAL MEASUREMENT

Charles T. Myers

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Charles T. Myers
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

This Memorandum is for interoffice use. It is not to be cited as a published report without the specific permission of the author.

Educational Testing Service
Princeton, New Jersey
September 1973

TM006 749

Bias and Interpretation: Cases for Ordinal Measurement

Charles T. Myers

According to Table No. 76 of the Statistical Abstract of the United States 1971 there were in this country during 1968 a total of 18,651 deaths due to accidental falls. Surely there is some truth in the rumor that there will soon be a demand for Congress to act promptly to repeal the law of gravitation. No other law works so constantly to limit movements and to diminish freedom. Furthermore, this law is obviously biased against obese people--people who already suffer from our social prejudices. No other law has been in effect for a longer time without having been revised to meet the needs of a changing society. If you detect in these remarks a note of sarcasm, perhaps it will suggest to you a certain skepticism about demands for educational reform. Freedom, equality, and progress are all greatly desired, but they may in fact conflict with each other; and not every demand or expressed need would in fact work to the benefit of society.

On the other hand, educational institutions may in fact suffer from undue rigidity and undue failure to examine the assumptions on which they were founded. It is the purpose of this paper to examine possible sources of bias in two of the recurrent topics of student dissatisfaction: the treatment of standardized test scores and the nature of course grades in college and high school. It has been common with respect to both of these forms of measurement to assume that they are properly treated as interval scales. This paper will discuss the alternative possibility that they should be treated as ordinal scales.

Among the objections that have been raised to conventional course grading systems, perhaps the most compelling are these: that the need to maintain a

good grade-point average inhibits students from taking chances with difficult courses or courses in areas with which they are unfamiliar, and that, similarly, the need to maintain a good grade-point average may corrode the relationship between teacher and student. So much is thought to depend upon the academic record and on grade-point averages (which are almost always the mean grade) that there is strong motivation for students to play either a subservient role to that of the teacher or a competitive role with regard to other students.

However, this situation is altered substantially if grades are assumed to be ordinal data rather than interval data, for in that case the appropriate average then becomes the median. In such a system a student need no longer fear a difficult or an unfamiliar field, nor an unfair teacher, because the one lowest grade out of four or five grades will have no effect whatsoever on that average. When grades are treated as ordinal scales, the students receive much of the benefit of "pass-fail" systems. But note that pass-fail grades may lower a student's average if he fails, while they cannot raise the average with a pass no matter how well the student has performed. The "pass-no-record" system is somewhat better than "pass-fail" in this regard, but it still lacks some of the virtues of ordinal-scale grading.

These comments on ordinal-scale grading are strategic rather than philosophical. On that aspect of the problem I will only ask the question, how would one establish the contention that grades should be treated as interval-scale data, particularly in colleges where the ability distribution is likely to have been sharply curtailed in the competitive admissions process? The use of systems of grading based on the assumption of a normal curve within each class clearly does not answer the problem.

Current theories of educational and psychological measurement have been derived in large measure from the statistics of gambling and anthropometric measurement. In both these fields it is common for distributions of events or of individuals to appear to approach normality. Test statistics also frequently result in distributions that are symmetrical and not dramatically nonnormal, although it is probably true that a majority of score distributions tend to be more platykurtic than normal. Perhaps because of this history and perhaps because the theory of linear regression has been more thoroughly developed than other alternatives, most discussions and investigations of tests assume normality for both univariate and multivariate distributions and linear relationships between all variables. Significance tests for linearity do not appear in the majority of reports of educational and psychological studies.

Perhaps these observations may be useful with respect to a related problem now being faced in the academic world--the problem of potential bias in the prediction of grades by the use of achievement and aptitude tests. In this case it may be particularly important to consider carefully the basic assumptions which have been applied to the treatment of test score data.

Validity studies such as those of Cleary (1968) and analyses such as those of Thorndike (1971) and Darlington (1971) use the analysis of covariance model and its assumptions regarding interval measurement and linear relationships. Test score distributions are descriptions of the relationship between a test composed of a set of items and a sample composed of a set of people. It can be demonstrated that a change in either set is likely to change the shape of the distribution. When tests are designed to maximize reliability in the classical test-theory model for a specified sample of people, the mean

score will approach the middle of the possible-score range and the distribution will usually tend to be symmetric. However, when either the test items or people are markedly changed so that the items are more difficult or the people less successful, then the score distribution tends to be skewed, particularly so when the mean score approaches the lower limit of possible scores.

In order to simplify the presentation of the analysis of covariance model it is common to represent score and criterion relationships by either a regression line or by an ellipse that represents contour lines of an idealized bivariate normal distribution. The center of the ellipse is located at the coordinates of the means of the two univariate distributions and the regression line passes through the center with a slope less than one (when, as is usual, scores scales are drawn with equal standard deviations). Another line through the center with a slope of one may be drawn to represent the major diagonal of the ellipse. This line represents, among other things, the equipercentile relationship between the test and the criterion--the line of relationship commonly used to describe the ordinal relationship between two variables. (See Figure 1.)

Insert Figure 1 about here

Now the point of this discussion is that the equipercentile line of relationship can never be identical with the regression line since tests are never perfectly reliable. Hence, if we are concerned with the relationship between two variables for two distinct groups with different mean scores on either variable, then the interval-scale interpretation of the relationship

must differ from the ordinal scale interpretation. Furthermore, in general, the ordinal interpretation will tend to be more favorable to the low scoring group.

The regression line and the equipercentile line of relationship between a predictor test and a predicted criterion cross at the mean and median when the bivariate distribution is linear and symmetrical. For higher scores than the mean, the regression line will be below the equipercentile line and the predicted criterion score will be lower than the equivalent equipercentile score. For scores below the mean, the relationship is reversed. If the bivariate distribution is irregular, skewed, or curvilinear, the two lines may not cross exactly at the mean, but the same general relationship will still exist. If one is concerned with two distinguishable groups that are being compared by means of the same predictor and criterion, then either one of two things must occur. Either the regression line for the low scoring group will be below the regression line for the high scoring group or the equipercentile line for the low scoring group will be above the other equipercentile line. Hence it is of considerable advantage to high scoring members of the low scoring group to have the equipercentile line used.

It should be noted that the use of the equipercentile line is, in the linear case, the same strategy as Thorndike has described as Case C and Darlington has described as Definition 2. However, it seems worthwhile to point out that these solutions may not only result in a "fairer" interpretation of scores in some sense, they may also avoid the use of assumptions that are perhaps sometimes difficult to justify or which may in fact not be justifiable at all.

Nominal scales are properly summarized by the mode and number of categories. Ordinal scales are properly summarized by the median and the semi-interquartile range: the relationship between two ordinal scales is described properly by equipercentile procedures and by rank correlation methods. Interval scales are properly summarized by the mean and standard deviation or by the mean and variance; and the relationship between two or more such scales is properly described by product-moment correlations and in some circumstances by analysis of covariance. Studies of bias should validate, insofar as possible, the assumptions that have been made in the model. When there is reason to doubt the linear model, it would be best to examine the relationships that would obtain for ordinal interpretations of the data. In general, the ordinal interpretation will be more favorable to low scoring groups than is an interpretation based on the assumption of an interval scale.

References

Cleary, T. A. Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Darlington, R. B. Another look at cultural fairness. Journal of Educational Measurement, 1971, 8, 71-82.

Thorndike, R. L. Concepts of culture-fairness. Journal of Educational Measurement, 1971, 8, 63-70.

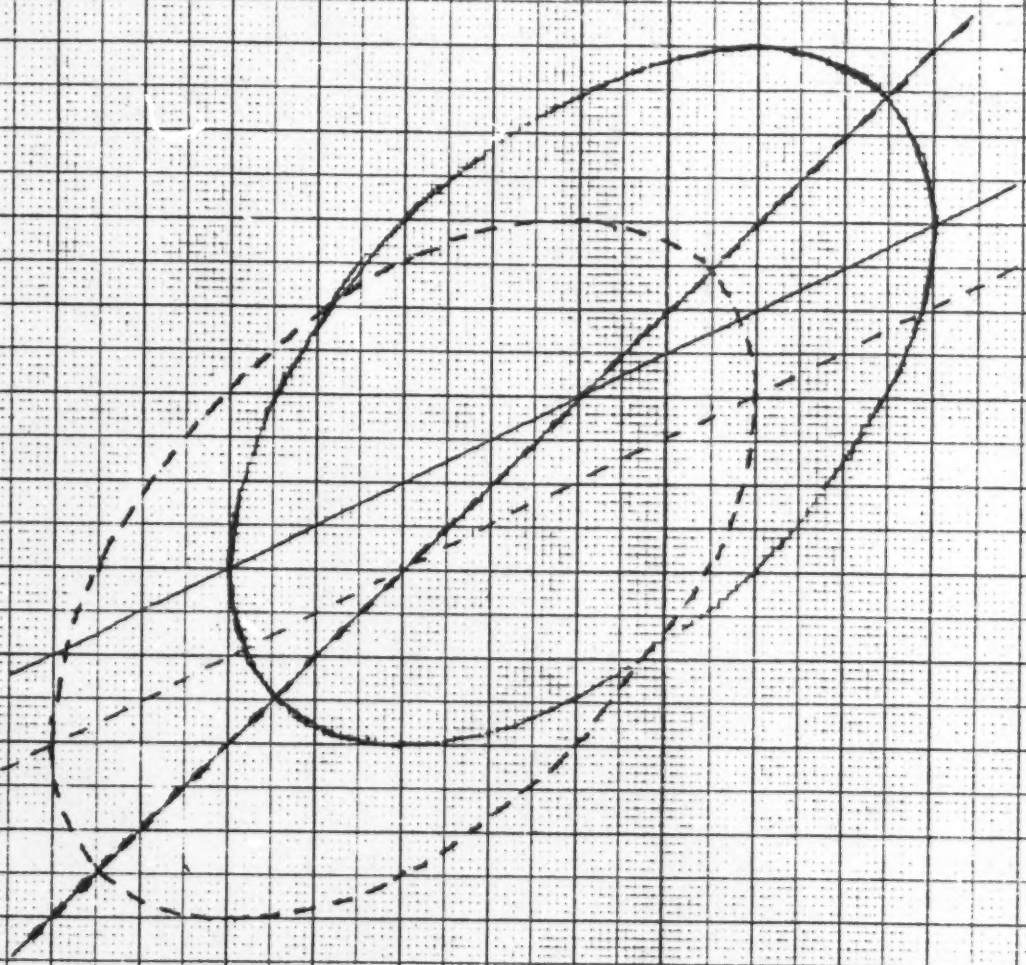

Figure 1

Hypothetical normal bivariate distributions for two groups with equal variances and equal correlations between variables and having a common equipercentile line, but having unequal means and regression line intercepts. Elliptical contour lines drawn at two standard deviations from the respective means. Difference between means equal to one standard deviation for either group. Correlations equal 0.50.

———————— Regression line, higher group

– – – – – – – – Regression line, lower group

— — — — — — — Equipercentile line, common to both groups